

1 CLAIMS:

- 2 1. A method for including a document in an index in a hyperlinked
3 environment, comprising the acts of:
4 receiving a document to be processed;
5 locating a set of documents that include hyperlinks to the
6 document;
7 retrieving anchortext associated with at least one of the
8 hyperlinks;
9 parsing the anchortext into one or more tokens;
10 for each token:
11 determining a weight for the token,
12 determining whether the weight assigned to the token exceeds a
13 threshold token weight; and
14 indexing the document under the token, if the token weight
15 assigned to the token exceeds the threshold token weight.
16
- 1
- 2 2. The method of claim 1, wherein the indexing act comprises including
3 in the index an indication of weight for each token under which each page is
4 indexed.
5

5

1 3. The method of claim 1, wherein the indexing act comprises assigning
2 to the token a location within the index corresponding to part of the page
3 being indexed that is allocated for tokens having a higher degree of
4 importance than other tokens in the same page.

1

2 4. The method of claim 3, wherein the indexing act comprises assigning
3 to the token a location within the index that corresponds to the beginning of
4 the page being indexed.

5

1 5. The method of claim 1, wherein the weight of each token is based on
2 its frequency of occurrence within the index.

1 6. The method of 1, wherein the act of determining a weight comprises:
2 determining a first frequency at which the anchortext appears in the
3 index;

4 determining a second frequency at which each token derived from the
5 anchortext appears in the index; and

6 assigning a weight to the token, wherein the weight is a function of the
7 first and second frequencies.

1

2 7. The method of claim 6, further comprising dividing the first frequency
3 by the second frequency to produce a weight quotient; and

4 multiplying the weight quotient by an anchor text count for the token.

5

1

2 8. The method of 6 further comprising determining a normalized weight
3 for each token.

4

4

1 9. A program product embedded in a machine-readable medium for
2 including a document in an index in a hyperlinked environment, comprising
3 the instructions for:

4 receiving a document to be processed;

5 locating a set of documents that include hyperlinks to the document;

6 retrieving anchortext associated with each hyperlink;

7 parsing the anchortext into one or more tokens;

8 and program instructions for each token comprising instructions for:

9 determining a weight for the token,

10 determining whether the weight assigned to the token exceeds a
11 threshold token weight; and

12 indexing the document under the token, if the token weight assigned to
13 the token exceeds the threshold token weight.

14

15

15

1 10. The computer program product of claim 9 wherein the indexing
2 instruction comprises including in the index an indication of weight for each
3 token under which each page is indexed.

4

1 11. The computer program product of claim 9, wherein the weight of each
2 token is based on its frequency of occurrence within the index.

1 12. The computer program product of claim 9, wherein the indexing act
2 comprises assigning to the token a location within the index that corresponds
3 to the beginning of the page being indexed.

13. The computer program product of claim 9, wherein the weight of each
5 token is based on its frequency of occurrence within the index.

14. The computer program product of claim 9, wherein the instruction of
determining a weight comprises:

determining a first frequency at which the anchortext appears in the
index;

10 determining a second frequency at which each token derived from the
anchortext appears in the index; and

assigning a weight to the token, wherein the weight is a function of the
first and second frequencies.

15 15. The program product of 13, further comprising the instruction of
determining a normalized weight for each token.

16. A system for indexing a document in a hyperlinked environment, comprising:

a receiver for receiving a document to be processed for inclusion in an index of documents;

5 a module for locating a set of documents that include hyperlinks to the document;

a module for retrieving anchortext associated with each hyperlink;

a parsing module for parsing the anchortext into one or more tokens;

and

10 a module for:

determining a weight for the token,

determining whether the weight assigned to the token exceeds a threshold token weight; and

15 indexing the document under the token, if the token weight assigned to the token exceeds the threshold token weight.